PhaseMax: Convex Phase Retrieval via Basis Pursuit

Tom Goldstein and Christoph Studer

Abstract

We consider the recovery of a (real- or complex-valued) signal from magnitude-only measurements, known as phase retrieval. We formulate phase retrieval as a convex optimization problem, which we call PhaseMax. Unlike other convex methods that use semidefinite relaxation and lift the phase retrieval problem to a higher dimension, PhaseMax operates in the original signal dimension. We show that the dual problem to PhaseMax is Basis Pursuit, which implies that phase retrieval can be performed using algorithms initially designed for sparse signal recovery. We develop sharp lower bounds on the success probability of PhaseMax for a broad range of random measurement ensembles, and we analyze the impact of measurement noise on the solution accuracy. We use numerical results to demonstrate the accuracy of our recovery guarantees, and we showcase the efficacy and limits of PhaseMax in practice.

I. INTRODUCTION

Phase retrieval is concerned with the recovery of an n-dimensional signal $\mathbf{x}^0 \in \mathcal{H}^n$, with $\mathcal{H}$ either $\mathbb{R}$ or $\mathbb{C}$, from $m \geq n$ squared-magnitude, noisy measurements [1]

$$b_i^2 = |\langle \mathbf{a}_i, \mathbf{x}^0 \rangle|^2 + \eta_i, \quad i = 1, 2, \ldots, m, \tag{1}$$

where $\mathbf{a}_i \in \mathcal{H}^n$, $i = 1, 2, \ldots, m$, are the (known) measurement vectors and $\eta_i \in \mathbb{R}$, $i = 1, 2, \ldots, m$, models measurement noise. Let $\hat{\mathbf{x}} \in \mathcal{H}^n$ be an approximation vector[1] to the true signal $\mathbf{x}^0$. We recover the signal $\mathbf{x}^0$ by solving the following convex problem called *PhaseMax*:

$$\text{(PM)} \quad \begin{cases} \underset{\mathbf{x} \in \mathcal{H}^n}{\text{maximize}} & \langle \mathbf{x}, \hat{\mathbf{x}} \rangle_{\Re} \\ \text{subject to} & |\langle \mathbf{a}_i, \mathbf{x} \rangle| \leq b_i, \quad i = 1, 2, \ldots, m. \end{cases}$$

T. Goldstein is with the Department of Computer Science, University of Maryland, College Park, MD (e-mail: tomg@cs.umd.edu).

C. Studer is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY (e-mail: studer@cornell.edu).

[1]Approximation vectors can be obtained via a variety of algorithms; see Section VI for the details.

Here, $\langle \mathbf{x}, \hat{\mathbf{x}} \rangle_{\Re}$ denotes the real-part of the inner product between the vectors $\mathbf{x}$ and $\hat{\mathbf{x}}$. The main idea behind PhaseMax is to find the vector $\mathbf{x}$ that is most aligned with the approximation vector $\hat{\mathbf{x}}$ and satisfies a convex relaxation of the measurement constraints in (1).

Our main goal is to develop sharp lower bounds on the probability with which PhaseMax succeeds in recovering the true signal $\mathbf{x}^0$, up to an arbitrary phase ambiguity that does not affect the measurement constraints in (1). By assuming noiseless measurements, one of our main results is as follows.

Theorem 1. *Consider the case of recovering a complex-valued signal $\mathbf{x} \in \mathbb{C}^n$ from m noiseless measurements of the form (1) with measurement vectors $\mathbf{a}_i$, $i = 1, 2, \ldots, m$, sampled independently and uniformly from the unit sphere. Let*

$$\mathrm{angle}(\mathbf{x}^0, \hat{\mathbf{x}}) = \arccos\left(\frac{\langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle_{\Re}}{\|\mathbf{x}^0\|_2 \|\hat{\mathbf{x}}\|_2} \right)$$

be the angle between the true vector $\mathbf{x}^0$ and the approximation $\hat{\mathbf{x}}$, and define the constant

$$\alpha = 1 - \tfrac{2}{\pi} \mathrm{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$$

that measures the approximation accuracy. Then, the probability that PhaseMax recovers the true signal $\mathbf{x}^0$, denoted by $p_{\mathbb{C}}(m, n)$, is bounded from below as follows:

$$p_{\mathbb{C}}(m, n) \geq 1 - \exp\left(-\frac{(\alpha(m-1) - 4n + 2)^2}{2(m-1)} \right) \tag{2}$$

whenever $\alpha(m-1) > 4n - 2$.

In words, if $m > (4n - 2)/\alpha + 1$ and $\alpha > 0$, then PhaseMax will succeed with non-zero probability. Furthermore, for a fixed signal dimension n and an arbitrary approximation vector $\hat{\mathbf{x}}$ that satisfies $\mathrm{angle}(\mathbf{x}^0, \hat{\mathbf{x}}) < \frac{\pi}{2}$, i.e., one that is not orthogonal to the vector $\mathbf{x}^0$, we can make the success probability of PhaseMax arbitrarily close to one by increasing the number of measurements m. As we shall see, our recovery guarantees are sharp and accurately predict the performance of PhaseMax in practice.

A. Convex Phase Retrieval via Basis Pursuit

It is quite intriguing that the following weighted Basis Pursuit problem [2], [3]

$$(\text{BP}) \quad \begin{cases} \underset{\mathbf{z} \in \mathcal{H}^m}{\text{minimize}} & \|\mathbf{B}\mathbf{z}\|_1 \\ \text{subject to} & \hat{\mathbf{x}} = \mathbf{A}\mathbf{z}, \end{cases}$$

with $\mathbf{B} = \mathrm{diag}(b_1, b_2, \ldots, b_m)$ and $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m]$ is the dual problem to (PM); see, e.g., [4, Lem. 1]. As a consequence, if PhaseMax succeeds, then the phases of the solution vector $\mathbf{z} \in \mathcal{H}^m$ to

(BP) are exactly the phases that were lost in the measurement process in (1), i.e., we have

$$y_i = \text{phase}(z_i) b_i = \langle \mathbf{a}_i, \mathbf{x}^0 \rangle, \quad i = 1, 2, \ldots, m,$$

with $\text{phase}(z) = z/|z|$ for $z \neq 0$ and $\text{phase}(0) = 1$. This observation not only reveals a fundamental connection between phase retrieval and sparse signal recovery, but also implies that Basis Pursuit solvers can be used to recover the signal from the phase-less measurements in (1).

B. Relevant Prior Art

Phase retrieval is a well-studied problem with a long history [5], [6] and enjoys widespread use in applications such as X-ray crystallography [7]–[9], microscopy [10], [11], imaging [12], and many more [13]–[16]. Early algorithms, such as the Gerchberg-Saxton [5] or Fienup [6] algorithms, rely on alternating projection to recover complex-valued signals from magnitude-only measurements. The papers [1], [17], [18] sparked new interest in the phase retrieval problem by showing that it can be relaxed to a semidefinite program. Prominent instances are PhaseLift [1] and PhaseCut [19]. These methods come with recovery guarantees but require the problem to be lifted to a higher dimensional space, which prevents their use for large-scale problems. More recently, a number of *non-convex* algorithms have been proposed (see e.g., [20]–[24]) that directly operate in the original signal dimension and exhibit excellent empirical performance. The algorithms in [20], [22]–[24] come with recovery guarantees that mainly rely on accurate initializers, such as the (truncated) spectral initializer [20], [23], the Null initializer [25], or the orthogonality-promoting method [24] (see Section VI for additional details). These initializers enable non-convex phase retrieval algorithms to succeed, given a sufficiently large number of measurements; see [26] for more details on the geometry of such non-convex problems.

C. Contributions and Paper Outline

In contrast to algorithms relying on semidefinite relaxation or non-convex problem formulations, we propose *PhaseMax*, a novel, convex method for phase retrieval that directly operates in the original signal dimension. In Section II, we establish a deterministic condition that guarantees uniqueness of the solution to the (PM) problem. Using this condition, we borrow methods from geometric probability in Section III in order to derive sharp lower bounds on the success probability for real- and complex-valued systems. Section V generalizes our results to a broader range of random measurement ensembles and to systems with measurement noise. We show in Section VI that randomly chosen approximation vectors are sufficient to ensure faithful recovery. We numerically demonstrate the sharpness of our recovery guarantees and showcases the practical limits of PhaseMax in Section VII. We conclude in Section VIII.

D. Notation

Lowercase and uppercase boldface letters stand for column vectors and matrices, respectively. For a complex-valued matrix $\mathbf{A}$, we denote its transpose and Hermitian transpose by $\mathbf{A}^T$ and $\mathbf{A}^*$, respectively; the real and imaginary parts are $\mathbf{A}_\Re$ and $\mathbf{A}_\Im$. The ith column of the matrix $\mathbf{A}$ is denoted by $\mathbf{a}_i$ and the kth entry of the ith vector $\mathbf{a}_i$ is $[\mathbf{a}_i]_k$; for a vector $\mathbf{a}$ without index, we simply denote the kth entry by a_k. We define the inner product between two complex-valued vectors $\mathbf{a}$ and $\mathbf{b}$ as $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^* \mathbf{b}$. We use j to denote the imaginary unit. The ℓ_2-norm and ℓ_1-norm of the vector $\mathbf{a}$ are $\|\mathbf{a}\|_2$ and $\|\mathbf{a}\|_1$, respectively.

II. UNIQUENESS CONDITION

There exist infinitely many vectors that satisfy the measurement constraints in (1). If $\mathbf{x}$ is a vector that satisfies (1), then any vector $\mathbf{x}' = e^{j\phi}\mathbf{x}$ for $\phi \in [0, 2\pi)$ also satisfies the constraints. In contrast, if $\mathbf{x}$ is a solution to (PM), then $e^{j\phi}\mathbf{x}$ with $\phi \neq 0$ will *not* be another solution. In fact, consider any vector $\mathbf{x}$ in the feasible set of (PM) with $\langle \mathbf{x}, \hat{\mathbf{x}} \rangle_\Im \neq 0$. By choosing $\omega = \text{phase}(\langle \mathbf{x}, \hat{\mathbf{x}} \rangle)$, we have

$$\langle \omega^* \mathbf{x}, \hat{\mathbf{x}} \rangle_\Re = |\langle \mathbf{x}, \hat{\mathbf{x}} \rangle| > \langle \mathbf{x}, \hat{\mathbf{x}} \rangle_\Re,$$

which implies that given such a vector $\mathbf{x}$, one can always increase the objective function of (PM) simply by *aligning* $\mathbf{x}$ with the approximation $\hat{\mathbf{x}}$ (i.e., modifying its phase so that $\langle \mathbf{x}, \hat{\mathbf{x}} \rangle$ is real valued). The following definition makes this observation rigorous.

Definition 1. *A vector $\mathbf{x}$ is said to be* aligned *with another vector $\hat{\mathbf{x}}$, if the inner product $\langle \mathbf{x}, \hat{\mathbf{x}} \rangle$ is real-valued and non-negative.*

From all the vectors that satisfy the measurement constraints in (1), there is only one that is a candidate solution to the convex problem (PM), which is also the solution that is aligned with $\hat{\mathbf{x}}$. For this reason, we adopt the following important convention throughout the rest of this paper.

The true vector $\mathbf{x}^0$ denotes a solution to (1) that is aligned with the approximation vector $\hat{\mathbf{x}}$.

Remark 1. *There is an interesting relation between the convex formulation of PhaseMax and the semidefinite relaxation method PhaseLift [1], [17], [18]. Recall that the set of solutions to any convex problem is always convex. However, the solution set of the measurement constraints (1) is invariant under phase rotations, and thus non-convex. It is therefore impossible to design a convex problem that yields this set of solutions. PhaseMax and PhaseLift differ in how they remove the phase ambiguity from the*

problem to enable a convex formulation. Rather than trying to identify the true vector $\mathbf{x}^0$, PhaseLift

reformulates the problem in terms of the quantity $\mathbf{x}^0(\mathbf{x}^0)^H$, which is unaffected by phase rotations in $\mathbf{x}^0$.

Hence, PhaseLift removes the rotation symmetry from the solution set, thus yielding a problem with a

convex set of solutions. PhaseMax does something much simpler: it pins down *the phase of the solution*

to an arbitrary quantity, thus removing the phase ambiguity and restoring convexity to the solution set.

This arbitrary phase choice is made when selecting the phase of the approximation $\hat{\mathbf{x}}$.

We are now ready to state a deterministic condition under which PhaseMax succeeds in recovering

the true vector $\mathbf{x}^0$. The result applies to the noiseless case, i.e., $\eta_i = 0$, $i = 1, 2, \ldots, m$. In this case, all

inequality constraints in (PM) are active at $\mathbf{x}^0$. The noisy case will be discussed in Section V-B.

Theorem 2. *The true vector $\mathbf{x}^0$ is the unique maximizer of (PM) if, for any unit vector $\boldsymbol{\delta} \in \mathcal{H}^n$ that is*

aligned with the approximation $\hat{\mathbf{x}}$,

$$\exists i, [\langle \mathbf{a}_i, \mathbf{x}^0 \rangle^* \langle \mathbf{a}_i, \boldsymbol{\delta} \rangle]_\Re > 0.$$

Proof. Suppose the conditions of this theorem hold, and consider some candidate solution $\mathbf{x}'$ in the

feasible set for (PM) with $\langle \mathbf{x}', \hat{\mathbf{x}} \rangle \geq \langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle$. Without loss of generality, we assume $\mathbf{x}'$ to be aligned

with $\hat{\mathbf{x}}$. Then, the vector $\boldsymbol{\Delta} = \mathbf{x}' - \mathbf{x}^0$ is also aligned with $\hat{\mathbf{x}}$, and satisfies

$$\langle \boldsymbol{\Delta}, \hat{\mathbf{x}} \rangle = \langle \mathbf{x}', \hat{\mathbf{x}} \rangle - \langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle \geq 0.$$

Since $\mathbf{x}'$ is a feasible solution for (PM), we have

$$|\langle \mathbf{a}_i, \mathbf{x}^0 + \boldsymbol{\Delta} \rangle|^2 = |\langle \mathbf{a}_i, \mathbf{x}^0 \rangle|^2 + 2[\langle \mathbf{a}_i, \mathbf{x}^0 \rangle^* \langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle]_\Re + |\langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle|^2 \leq b_i^2, \ \forall i.$$

But $|\mathbf{a}_i^T \mathbf{x}^0|^2 = b_i^2$, and so

$$[\langle \mathbf{a}_i, \mathbf{x}^0 \rangle^* \langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle]_\Re \leq -\tfrac{1}{2}|\langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle|^2 \leq 0, \ \forall i.$$

Now, if $\|\boldsymbol{\Delta}\|_2 > 0$, then the unit-length vector $\boldsymbol{\delta} = \boldsymbol{\Delta}/\|\boldsymbol{\Delta}\|_2$ satisfies $[\langle \mathbf{a}_i, \mathbf{x}^0 \rangle^* \langle \mathbf{a}_i, \boldsymbol{\delta} \rangle]_\Re \leq 0$ for all i,

which contradicts the hypothesis of the theorem. It follows that $\|\boldsymbol{\Delta}\|_2 = 0$ and $\mathbf{x}' = \mathbf{x}^0$. $\square$

Theorem 2 has an intuitive geometrical interpretation. If $\mathbf{x}^0$ is an optimal point and $\boldsymbol{\delta}$ is an ascent

direction, then one cannot move in the direction of $\boldsymbol{\delta}$ starting at $\mathbf{x}^0$ without leaving the feasible set. This

condition is met if there is an $\mathbf{a}_i$ such that $\mathbf{x}^0$ and $\boldsymbol{\delta}$ both lie on the same side of the plane through the

origin orthogonal to the measurement vector $\mathbf{a}_i$.

III. Preliminaries: classical sphere covering problems and geometric probability

In order to derive sharp conditions on the success probability of PhaseMax, we require a set of tools from geometric probability. Many classical problems in geometric probability involve calculating the likelihood of a sphere being covered by random "caps," or semi-spheres, which we define below.

Definition 2. *Consider the set $\mathcal{S}_{\mathcal{H}}^{n-1} = \{\mathbf{x} \in \mathcal{H}^n \,|\, \|\mathbf{x}\|_2 = 1\}$, the unit sphere embedded in $\mathcal{H}^n$. Given a vector $\mathbf{a} \in \mathcal{H}^n$, the cap centered at $\mathbf{a}$ with central angle θ is defined as*

$$\mathcal{C}_{\mathcal{H}}(\mathbf{a}, \theta) = \{\boldsymbol{\delta} \in \mathcal{S}_{\mathcal{H}}^{n-1} \,|\, \langle \mathbf{a}, \boldsymbol{\delta} \rangle_{\Re} > \cos(\theta)\}. \tag{3}$$

This cap contains all vectors that form an angle with $\mathbf{a}$ of less than θ radians. When $\theta = \pi/2$, we have a semisphere centered at $\mathbf{a}$, which is simply denoted by

$$\mathcal{C}_{\mathcal{H}}(\mathbf{a}) = \mathcal{C}_{\mathcal{H}}(\mathbf{a}, \pi/2) = \{\boldsymbol{\delta} \in \mathcal{S}_{\mathcal{H}}^{n-1} \,|\, \langle \mathbf{a}, \boldsymbol{\delta} \rangle_{\Re} > 0\}. \tag{4}$$

We say that a collection of caps *covers* the entire sphere if the sphere is contained in the union of the caps. Before we can say anything useful about when a collection of caps covers the sphere, we will need the following classical result, which is often attributed to Schläfli [27]. Proofs that use simple induction methods can be found in [28]–[30].

Lemma 1. *Consider a sphere $\mathcal{S}_{\mathbb{R}}^{n-1} \subset \mathbb{R}^n$. Suppose we slice the sphere with k unique planes through the origin. These planes divide the sphere into*

$$r(n, k) = 2 \sum_{i=0}^{n-1} \binom{k-1}{i}$$

regions.

Classical results in geometric probability study the likelihood of a sphere being covered by random caps with centers chosen independently and uniformly from the sphere's surface. For our purposes, we need to study the more specific case in which caps are only chosen from a subset of the sphere. While calculating this probability is hard in general, it is quite simple when the set obeys the following symmetry condition.

Definition 3. *We say that the set $\mathcal{A}$ is* symmetric *if, for all $\mathbf{x} \in \mathcal{A}$, we also have $-\mathbf{x} \in \mathcal{A}$.*

We are now ready to prove fairly general results that state when the sphere is covered by random caps.

Lemma 2. *Consider some non-empty symmetric set $\mathcal{A} \subset \mathcal{S}_{\mathbb{R}}^{n-1}$. Choose some set of $m_{\mathcal{A}}$ measure-*

ments $\{\mathbf{a}_i\}_{i=1}^{m_{\mathcal{A}}}$ uniformly from $\mathcal{A}$. Then, the caps $\{\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i)\}$ cover the sphere $\mathcal{S}_{\mathbb{R}}^{n-1}$ with probability

$$p_{cover}(m_{\mathcal{A}}, n) = 1 - \frac{1}{2^{m_{\mathcal{A}}-1}} \sum_{k=0}^{n-1} \binom{m_{\mathcal{A}} - 1}{k}.$$

This is the probability of turning up n or more heads when flipping $m_{\mathcal{A}} - 1$ fair coins.

Proof. Consider the following two-step process for constructing the set $\{\mathbf{a}_i\}$. First, we sample $m_{\mathcal{A}}$ vectors $\{\mathbf{a}_i'\}$ independently and uniformly from $\mathcal{A}$. Second, we define $\mathbf{a}_i = c_i \mathbf{a}_i'$, where $\{c_i\}$ are i.i.d. Bernoulli variables that take value $+1$ or -1 with probability $\frac{1}{2}$. We can think of this second step as randomly "flipping" a subset of uniform random vectors. Since $\mathcal{A}$ is symmetric and $\{\mathbf{a}_i'\}$ is sampled independently and uniformly, the random vectors $\{\mathbf{a}_i\}$ also have an independent and uniform distribution over $\mathcal{A}$. This construction may seem superfluous since both $\{\mathbf{a}_i\}$ and $\{\mathbf{a}_i'\}$ have the same distribution, but we will see below that this becomes useful.

Given a particular set of coin flips $\{c_i\}$, we can write the set of points that are *not* covered by the caps $\{\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i)\}$ as

$$\bigcap_i \mathcal{C}_{\mathbb{R}}(-\mathbf{a}_i) = \bigcap_i \mathcal{C}_{\mathbb{R}}(-c_i \mathbf{a}_i'). \tag{5}$$

Note that there are $2^{m_{\mathcal{A}}}$ such intersections that can be formed, one for each choice of the sequence $\{c_i\}$. The caps $\{\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i)\}$ cover the sphere whenever the intersection (5) is empty. Consider the set of planes $\{\{\mathbf{x} \mid \langle \mathbf{a}_i', \mathbf{x} \rangle = 0\}\}$. From Lemma 1, we know that $m_{\mathcal{A}}$ planes with a common intersection point divide the sphere into

$$r(n, m_{\mathcal{A}}) = 2 \sum_{k=0}^{n-1} \binom{m_{\mathcal{A}} - 1}{k}$$

non-empty regions. Each of these regions corresponds to the intersection (5) for one possible choice of $\{c_i\}$. Therefore, of the $2^{m_{\mathcal{A}}}$ possible intersections, at most $r(n, m_{\mathcal{A}})$ of them are non-empty. Since the sequence $\{c_i\}$ is random, each intersection is equally likely to be chosen, and so the probability of covering the sphere is

$$p_{\text{cover}}(m_{\mathcal{A}}, n) = 1 - \frac{r(n, m_{\mathcal{A}})}{2^{m_{\mathcal{A}}}}.$$

$\square$

Remark 2. *Several papers have studied the probability of covering the sphere using points independently and uniformly chosen over the entire sphere. The only aspect that is unusual about Lemma 2 is the observation that this probability remains the same if we restrict our choices to the set $\mathcal{A}$, provided $\mathcal{A}$ is symmetric. We note that this result was observed by Gilbert [29] in the case $n = 3$, and we generalize it*

to any $n > 1$ using a similar argument.

We now present a somewhat more complicated covering theorem. The next result considers the case where the measurement vectors are drawn only from a semisphere. We consider the question of whether these vectors cover enough area to contain not only their home semisphere, but another nearby semisphere as well.

Lemma 3. *Consider two vectors $\mathbf{x}, \mathbf{y} \subset \mathcal{S}_{\mathbb{R}}^{n-1}$, and the caps $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$ and $\mathcal{C}_{\mathbb{R}}(\mathbf{y})$. Let $\alpha = 1 - \frac{2}{\pi} \operatorname{angle}(\mathbf{x}, \mathbf{y})$ be a measure of the similarity between the vectors $\mathbf{x}$ and $\mathbf{y}$. Draw some collection $\{\mathbf{a}_i \in \mathcal{C}_{\mathbb{R}}(\mathbf{x})\}_{i=1}^{m}$ of m vectors uniformly from $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$ so that $\alpha(m - 1) > 2n$. Then*

$$\mathcal{C}_{\mathbb{R}}(\mathbf{y}) \subset \bigcap_i \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i)$$

with probability at least

$$p_{cover}(m, n; \mathbf{x}, \mathbf{y}) \geq 1 - \exp\left(-\frac{(\alpha(m-1) - 2n)^2}{2(m-1)}\right).$$

Proof. Due to rotational symmetry, we assume $\mathbf{y} = [1, 0, \ldots, 0]^T$ without loss of generality. Consider the point $\tilde{\mathbf{x}} = [x_1, -x_2, \ldots, -x_n]^T$. This is the reflection of $\mathbf{x}$ over $\mathbf{y}$. Suppose we have some collection $\{\mathbf{a}_i\}$ independently and uniformly distributed on the entire sphere. Consider the collection of vectors

$$\mathbf{a}_i' = \begin{cases} \mathbf{a}_i, & \text{if } \langle \mathbf{a}_i, \mathbf{x} \rangle \geq 0 & (6) \\[2mm] \mathbf{a}_i - 2\langle \mathbf{a}_i, \mathbf{y} \rangle \mathbf{y} & \text{if } \langle \mathbf{a}_i, \mathbf{x} \rangle < 0, \langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle < 0 & (7) \\[2mm] -\mathbf{a}_i & \text{if } \langle \mathbf{a}_i, \mathbf{x} \rangle < 0, \langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle \geq 0. & (8) \end{cases}$$

The mapping $\mathbf{a}_i \to \mathbf{a}_i'$ maps the lower half sphere $\{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle < 0\}$ onto the upper half sphere $\{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle > 0\}$ using a combination of reflections and translations. Indeed, for all i we have $\langle \mathbf{a}_i', \mathbf{x} \rangle \geq 0$. This is clearly true in case (6) and (8). In case (7), observe that $\langle \mathbf{a}_i, \mathbf{x} \rangle + \langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle = 2[\mathbf{a}_i]_1 x_1$. We can then calculate

$$\langle \mathbf{a}_i', \mathbf{x} \rangle = \langle \mathbf{a}_i, \mathbf{x} \rangle - 2\langle \mathbf{a}_i, \mathbf{y} \rangle \langle \mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{a}_i, \mathbf{x} \rangle - 2[\mathbf{a}_i]_1 x_1 = -\langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle \geq 0.$$

Because the mapping $\mathbf{a}_i \to \mathbf{a}_i'$ is onto and (piecewise) isometric, $\{\mathbf{a}_i'\}$ will be uniformly distributed over the half sphere $\{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x}^0 \rangle > 0\}$ whenever $\{\mathbf{a}_i\}$ are independently and uniformly distributed over the entire sphere.

Consider the "hourglass" shaped, symmetric set

$$\mathcal{A} = \{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle \geq 0, \langle \mathbf{a}, \tilde{\mathbf{x}} \rangle \geq 0\} \cup \{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle \leq 0, \langle \mathbf{a}, \tilde{\mathbf{x}} \rangle \leq 0\}.$$

We now make the following claim: $\mathcal{C}_{\mathbb{R}}(\mathbf{y}) \subset \bigcap_i \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$ whenever

$$\mathcal{S}_{\mathbb{R}}^{n-1} \subset \bigcap_{\mathbf{a}_i \in \mathcal{A}} \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i). \tag{9}$$

In words, if the caps defined by the subset of $\{\mathbf{a}_i\}$ in $\mathcal{A}$ cover the entire sphere, then the caps $\{\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')\}$ (which have centers in $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$) not only cover $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$, but also cover its neighbor cap $\mathcal{C}_{\mathbb{R}}(\mathbf{y})$. To justify this claim, suppose that (9) holds. Choose some $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{R}}(\mathbf{y})$. This point is covered by some cap $\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i)$ with $\mathbf{a}_i \in \mathcal{A}$. If $\langle \mathbf{a}_i, \mathbf{x} \rangle \geq 0$, then $\mathbf{a}_i = \mathbf{a}_i'$ and $\boldsymbol{\delta}$ is covered by $\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$. If $\langle \mathbf{a}_i, \mathbf{x} \rangle < 0$, then

$$\langle \boldsymbol{\delta}, \mathbf{a}_i' \rangle = \langle \boldsymbol{\delta}, \mathbf{a}_i - 2\langle \mathbf{a}_i, \mathbf{y} \rangle \mathbf{y} \rangle = \langle \boldsymbol{\delta}, \mathbf{a}_i \rangle - 2 \langle \mathbf{a}_i, \mathbf{y} \rangle \langle \boldsymbol{\delta}, \mathbf{y} \rangle \geq \boldsymbol{\delta} \langle \mathbf{a}, \mathbf{a}_i \rangle \geq 0.$$

Note we have used the fact that $\langle \boldsymbol{\delta}, \mathbf{y} \rangle$ is real and non-negative because $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{R}}(y)$. We have also used $\langle \mathbf{a}_i, \mathbf{y} \rangle = [\mathbf{a}_i]_1 = \frac{1}{2}(\langle \mathbf{a}_i, \mathbf{x} \rangle + \langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle) < 0$, which follows from the definition of $\tilde{\mathbf{x}}$ and the definition of $\mathcal{A}$. Since $\boldsymbol{\delta} \langle \mathbf{a}, \mathbf{a}_i' \rangle \geq 0$, we have $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$, which proves our claim.

We can now see that the probability that $\mathcal{C}_{\mathbb{R}}(\mathbf{y}) \subset \bigcap_i \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$ is at least as high as the probability that (9) holds. Let $p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y} \mid m_{\mathcal{A}})$ denote the probability of covering $\mathcal{C}(\mathbf{y})$ conditioned on the number $m_{\mathcal{A}}$ of points lying in $\mathcal{A}$. From Lemma 2, we know that $p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y} \mid m_{\mathcal{A}}) \geq p_{\text{cover}}(m_{\mathcal{A}}, n)$. As noted in Lemma 2, this is the chance of turning up n or more heads when flipping $m_{\mathcal{A}} - 1$ fair coins.

The probability $p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y})$ is then given by

$$p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y}) = \mathbb{E}_{m_{\mathcal{A}}}[p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y} \mid m_{\mathcal{A}})] \geq \mathbb{E}_{m_{\mathcal{A}}}[p_{\text{cover}}(m_{\mathcal{A}}, n)].$$

The expression on the right hand side is the probability of getting n or more heads when one fair coin is flipped for every measurement $\mathbf{a}_i$ that lies in $\mathcal{A}$.

The region $\mathcal{A}$ is defined by two planes that intersect at an angle of $\beta = \text{angle}(\mathbf{x}, \tilde{\mathbf{x}}) = 2\,\text{angle}(\mathbf{x}, \mathbf{y})$. The probability of a random point $\mathbf{a}_i$ lying in $\mathcal{A}$ is given by $\alpha = \frac{2\pi - 2\beta}{2\pi} = 1 - \frac{2\beta}{\pi}$, which is the fraction of the unit sphere that lies either above or below both planes. The probability of a measurement $\mathbf{a}_i$ contributing to the heads count is half the probability of it lying in $\mathcal{A}$, or $\frac{1}{2}\alpha$. The probability of turning up n or more heads is therefore given by

$$1 - \sum_{k=0}^{n-1} \left(\frac{1}{2}\alpha\right)^k \left(1 - \frac{1}{2}\alpha\right)^{m-k-1} \binom{m-1}{k}.$$

Using Hoeffding's inequality, we obtain the following lower bound

$$p_{\text{cover}}(m, n) \geq 1 - \sum_{k=0}^{n-1} \left(\frac{1}{2}\alpha\right)^k \left(1 - \frac{1}{2}\alpha\right)^{m-k-1} \binom{m-1}{k} \geq 1 - \exp\left(\frac{-(\alpha(m-1) - 2n)^2}{2(m-1)}\right),$$

which is only valid for $\alpha(m-1) > 2n$. □

Remark 3. *In the proof of Lemma 3 we obtained $\mathbb{E}_{m_\mathcal{A}}[p_{cover}(m_\mathcal{A}, n)]$ using an intuitive argument about coin flipping probabilities. This expectation could have been obtained more rigorously (but with considerably more pain) using the method of probability generating functions.*

Lemma 3 contains most of the machinery needed for the proofs that follow. In the sequel, we prove a number of exact reconstruction theorems for (PM). Most of the results rely of short arguments followed by the invocation of Lemma 3.

We finally state a result that bounds the probability of covering the sphere with caps of small central angle from below. The following Lemma is a direct corollary of the results of Burgisser, Cucker, and Lotz in [31]. A derivation that uses their results is given in Appendix A.

Lemma 4. *Let $n \geq 9$, and $m > 2n$. Then the probability of covering the sphere $S_\mathbb{R}^{n-1}$ with independent uniformly sampled caps of central angle $\phi \leq \pi/2$ is lower bounded by*

$$p_{cover}(m, n, \phi) \geq 1 - \frac{(em)^n \sqrt{n-1}}{(2n)^{n-1}} \exp\left(-\frac{\sin^{n-1}(\phi)(m-n)}{\sqrt{8n}}\right)\cos(\phi) - \exp\left(-\frac{(m-2n+1)^2}{2m-2}\right).$$

IV. RECOVERY GUARANTEES

Using the uniqueness condition provided by Theorem 2 and the tools derived in Section III, we now develop sharp lower bounds on the success probability of PhaseMax for noiseless real- and complex-valued systems. The noisy case will be discussed in Section V-B.

A. The Real Case

We now study problem (PM) in the case that the unknown signal and measurement vectors are real valued. Consider some collection of measurement vectors $\{\mathbf{a}_i\}$ drawn independently and uniformly from $\mathcal{S}_\mathbb{R}^{n-1}$. For simplicity, we also consider the collection $\{\tilde{\mathbf{a}}_i\} = \{\text{phase}(\langle \mathbf{a}_i, \mathbf{x}^0 \rangle)\mathbf{a}_i\}$ of aligned vectors that satisfy $\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle \geq 0$ for all i. Using this notation, Theorem 2 can be rephrased as a simple geometric condition.

Corollary 1. *Consider the set $\{\tilde{\mathbf{a}}_i\} = \{\text{phase}(\langle \mathbf{a}_i, \mathbf{x}^0 \rangle)\mathbf{a}_i\}$ of aligned measurement vectors. Define the half sphere of aligned ascent directions*

$$\mathcal{D}_\mathbb{R} = \mathcal{C}_\mathbb{R}(\hat{\mathbf{x}}) = \{\boldsymbol{\delta} \in \mathcal{S}_\mathbb{R}^{n-1} | \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle \in \mathbb{R} \geq 0\}.$$

The true vector $\mathbf{x}^0$ will be the unique maximizer of (PM) if

$$\mathcal{D}_\mathbb{R} \subset \bigcap_i \mathcal{C}_\mathbb{R}(\tilde{\mathbf{a}}_i).$$

Proof. Choose some ascent direction $\boldsymbol{\delta} \in \mathcal{D}_\mathbb{R}$. If the assumptions of this Corollary hold, then there is some i with $\boldsymbol{\delta} \in \mathcal{C}_\mathbb{R}(\tilde{\mathbf{a}}_i)$, and so $\langle \tilde{\mathbf{a}}_i, \boldsymbol{\delta} \rangle \geq 0$. Since this is true for any $\boldsymbol{\delta} \in \mathcal{D}_\mathbb{R}$, the conditions of Theorem 2 are satisfied and exact reconstruction holds. $\square$

Using this observation, we can develop the following lower bound on the success probability of PhaseMax for real-valued systems.

Theorem 3. *Consider the case of recovering a real-valued signal $\mathbf{x} \in \mathbb{R}^n$ from m noiseless measurements of the form (1) with measurement vectors $\mathbf{a}_i$, $i = 1, 2, \ldots, m$, sampled independently and uniformly from the unit sphere $\mathcal{S}_\mathbb{R}^{n-1}$. Then, the probability that PhaseMax recovers the true signal $\mathbf{x}^0$, denoted by $p_\mathbb{R}(m, n)$, is bounded from below as follows:*

$$p_\mathbb{R}(m, n) \geq 1 - \exp\left(\frac{-(\alpha(m-1) - 2n)^2}{2(m-1)} \right),$$

where $\alpha = 1 - \frac{2}{\pi} \operatorname{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$ and $\alpha(m-1) > 2n$.

Proof. Consider the set of m independent and uniformly sampled measurements $\{\mathbf{a}_i \in \mathcal{S}_\mathbb{R}^{n-1}\}_{i=1}^m$. The aligned vectors $\{\tilde{\mathbf{a}}_i = \operatorname{phase}(\langle \mathbf{a}_i, \mathbf{x}^0 \rangle)\mathbf{a}_i\}$ are uniformly distributed over the half sphere $\mathcal{C}_\mathbb{R}(\mathbf{x}^0)$. Exact reconstruction happens when the condition in Corollary 1 holds. To obtain a lower bound on the probability of this occurrence, we can simply invoke Lemma 3 with $\mathbf{x} = \mathbf{x}^0$ and $\mathbf{y} = \hat{\mathbf{x}}$. $\square$

B. The Complex Case

We now prove Theorem 1 given in Section I, which characterizes the success probability of PhaseMax for phase retrieval in complex-valued systems. For clarity, we re-state our result in shorter form.

Theorem 1. *Consider the case of recovering a complex-valued signal $\mathbf{x}^0 \in \mathbb{C}^n$ from m noiseless measurements of the form (1), with $\{\mathbf{a}_i\}_{i=1}^m$ sampled independently and uniformly from the unit sphere $\mathcal{S}_\mathbb{C}^{n-1}$. Then, the probability that PhaseMax recovers the true signal $\mathbf{x}^0$ is bounded from below as follows:*

$$p_\mathbb{C}(m, n) \geq 1 - \exp\left(-\frac{(\alpha(m-1) - 4n + 2)^2}{2(m-1)} \right),$$

where $\alpha = 1 - \frac{2}{\pi} \operatorname{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$ and $\alpha(m-1) > 4n - 2$.

Proof. Consider the set $\{\tilde{\mathbf{a}}_i\} = \{\text{phase}(\langle \mathbf{a}_i, \mathbf{x}^0 \rangle)\mathbf{a}_i\}$ of aligned measurement vectors. Define the half sphere of aligned ascent directions

$$\mathcal{D}_{\mathbb{C}} = \{\boldsymbol{\delta} \in \mathcal{S}_{\mathbb{C}}^{n-1} \,|\, \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle_{\Re} \geq 0, \ \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle_{\Im} = 0\}.$$

By Corollary 1, the true signal $\mathbf{x}^0$ will be the unique maximizer of (PM) if

$$\mathcal{D}_{\mathbb{C}} \subset \bigcap_i \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i). \tag{10}$$

Let us bound the probability of this event. Consider the set $\mathcal{A} = \{\boldsymbol{\delta} \,|\, \langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} = 0\}$. We now claim that (10) holds whenever

$$\mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}}) \cap \mathcal{A} \subset \bigcap_i \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i). \tag{11}$$

To prove this claim, consider some $\boldsymbol{\delta} \in \mathcal{D}_{\mathbb{C}}$. To keep notation light, we will assume without loss of generality that $\|\mathbf{x}^0\|_2 = 1$. Form the vector $\boldsymbol{\delta}' = \boldsymbol{\delta} + j\langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} \mathbf{x}^0$, which is the projection of $\boldsymbol{\delta}$ onto $\mathcal{A}$. It is clear that $\boldsymbol{\delta}' \in \mathcal{A}$ because

$$\langle \boldsymbol{\delta}', \mathbf{x}^0 \rangle = \langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle + \langle j\langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} \mathbf{x}^0, \mathbf{x}^0 \rangle = \langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle - j\langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} \langle \mathbf{x}^0, \mathbf{x}^0 \rangle = \langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle - j\langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} = \langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Re},$$

which is real valued. Furthermore, $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}})$ because

$$\langle \boldsymbol{\delta}', \hat{\mathbf{x}} \rangle = \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle + \langle j\langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} \mathbf{x}^0, \hat{\mathbf{x}} \rangle = \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle - j\langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} \langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle.$$

The first term on the right is real-valued and non-negative (because $\boldsymbol{\delta} \in \mathcal{D}_{\mathbb{C}}$), and the second term is complex valued (because $\mathbf{x}^0$ is assumed to be aligned with $\hat{\mathbf{x}}$). It follows that $\langle \boldsymbol{\delta}', \hat{\mathbf{x}} \rangle_{\Re} \geq 0$ and $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}})$. Since we already showed that $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}})$, we have $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}}) \cap \mathcal{A}$. Suppose now that (11) holds. The claim will be proved if we can show that $\boldsymbol{\delta} \in \mathcal{D}$ is covered by one of the $\mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i)$. Since $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}}) \cap \mathcal{A}$, there is some i with $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i)$. But then

$$0 \leq \langle \boldsymbol{\delta}', \tilde{\mathbf{a}}_i \rangle_{\Re} = \langle \boldsymbol{\delta}, \tilde{\mathbf{a}}_i \rangle_{\Re} + \langle j\langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle_{\Im} \mathbf{x}^0, \tilde{\mathbf{a}}_i \rangle_{\Re} = \langle \boldsymbol{\delta}, \tilde{\mathbf{a}}_i \rangle_{\Re}. \tag{12}$$

We see that $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i)$, and the claim is proved.

We now know that exact reconstruction happens whenever condition (11) holds. We can put a bound on the frequency of this using Lemma 3. Note that the sphere $S_{\mathbb{C}}^{n-1}$ is isomorphic to $S_{\mathbb{R}}^{2n-1}$, and the set $\mathcal{A}$ is isomorphic to the sphere $S_{\mathbb{R}}^{2n-2}$. The aligned vectors $\{\tilde{\mathbf{a}}_i\}$ are uniformly distributed over a half sphere in $\mathcal{C}_{\mathbb{C}}(\mathbf{x}^0) \cap \mathcal{A}$, which is isomorphic to the upper half sphere in $S_{\mathbb{R}}^{2n-2}$. The probability of these vectors covering the cap $\mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}}) \cap \mathcal{A}$ is thus given by $p_{\text{cover}}(m, 2n-1; \mathbf{x}^0, \hat{\mathbf{x}})$ from Lemma 3. □

Remark 4. *Theorems 1 and 3 guarantee exact recovery for a sufficiently large number of measurements m provided that* $\mathrm{angle}(\mathbf{x}^0, \hat{\mathbf{x}}) < \frac{\pi}{2}$. *In the case* $\mathrm{angle}(\mathbf{x}^0, \hat{\mathbf{x}}) > \frac{\pi}{2}$, *our theorems guarantee convergence to* $-\mathbf{x}^0$ *(which is also a valid solution) for sufficiently large m. Our theorems only fail for large m if* $\arccos(\mathbf{x}^0, \hat{\mathbf{x}}) = \pi/2$, *which happens with probability zero when the approximation vector $\hat{\mathbf{x}}$ is generated at random. See Section VI for more details.*

V. GENERALIZATIONS

Our theory thus far addressed the idealistic case in which the measurement vectors are independently and uniformly sampled from a unit sphere and for noiseless measurements. We now extend our results to more general random measurement ensembles and to noisy measurements.

A. Generalized Measurement Ensembles

The theorems of Section IV require the measurement vectors $\{\mathbf{a}_i\}$ to be drawn independently and uniformly from the surface of the unit sphere. This condition can easily be generalized to other sampling ensembles. In particular, our results still hold for all *rotationally symmetric distributions*. A distribution D is rotationally symmetric if the distribution of $\mathbf{a}/\|\mathbf{a}\|_2$ is uniform over the sphere when $\mathbf{a} \sim D$. For such a distribution, one can make the change of variables $\mathbf{a} \leftarrow \mathbf{a}/\|\mathbf{a}\|_2$, and then apply Theorems 1 and 3 to the resulting problem. Note that this change of variables does not change the feasible set for (PM), and thus does not change the solution. Consequently, the same recovery guarantees apply to the original problem without explicitly implementing this change of variables. We thus have the following simple corollary.

Corollary 2. *The results of Theorem 1 and Theorem 3 still hold if the samples $\{\mathbf{a}_i\}$ are drawn from a multivariate Gaussian distribution with independent and identically distributed (i.i.d.) entries.*

Proof. A multivariate Gaussian distribution with i.i.d. entries is rotationally symmetric, and thus the change of variables $\mathbf{a} \leftarrow \mathbf{a}/\|\mathbf{a}\|_2$ yields an equivalent problem with measurements sampled uniformly from the unit sphere. $\qquad\square$

What happens when the distribution is not spherically symmetric? In this case, we can still guarantee recovery, but we require a larger number of measurements. The following result is, analogous to Theorem 1, for the noiseless complex case.

Theorem 4. *Suppose that m_D measurement vectors $\{\mathbf{a}_i\}_{i=1}^{m_D}$ are drawn from the unit sphere with (possibly non-uniform) probability density function $D : S_{\mathbb{C}}^{n-1} \to \mathbb{R}$. Let $\ell_D \leq \inf_{\mathbf{x} \in S_{\mathbb{C}}^{n-1}} D(\mathbf{x})$ be a lower bound on D over the unit sphere and let $\alpha = 1 - \frac{2}{\pi} \mathrm{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$ as above. We use $s_n = \frac{2\pi^n}{\Gamma(n)}$ to denote the*

"surface area" of the complex sphere $\mathcal{S}_{\mathbb{C}}^{n-1}$, and set $m_U = \lfloor m_D s_n \ell_D \rfloor$. Then, exact reconstruction is guaranteed with probability at least

$$1 - \exp\left(-\frac{(\alpha(m_U - 1) - 4n + 2)^2}{2m_D - 2} \right)$$

whenever $\alpha(m_U - 1) > 4n - 2$ and $\ell_D > 0$. In other words, exact recovery with m_D non-uniform measurements happens at least as often as with m_U uniform measurements.

Proof. We compare two measurement models, a uniform measurement model in which m_U measurements are drawn uniformly from a unit sphere, and a non-uniform measurement model in which m_D measurements are drawn from the distribution D. Note that the sphere $\mathcal{S}_{\mathbb{C}}^{n-1}$ has surface area $s_n = \frac{2\pi^n}{\Gamma(n)}$, and the uniform density function U on this sphere has constant value s_n^{-1}. Consider some collection of measurements $\{\mathbf{a}_i^U\}_{i=1}^{m_U}$ drawn from the uniform model. The joint probability density of this measurement ensemble is

$$m_U! s_n^{-m_U}.$$

Now consider some ensemble $\{\mathbf{a}_i^D\}_{i=1}^{m_D}$ drawn with density D. The event that $\{\mathbf{a}_i^U\} \subset \{\mathbf{a}_i^D\}$ has density

$$\frac{m_D!}{(m_D - m_U)!} \prod_{i=1}^{m_U} D(\mathbf{a}_i).$$

The ratio of the non-uniform density to the uniform density is

$$\binom{m_D}{m_U} \prod_{i=1}^{m_U} s_n D(\mathbf{a}_i) \geq \binom{m_D}{m_U} (s_n \ell_D)^{m_U} \geq \left(\frac{m_D s_n \ell_D}{m_U} \right)^{m_U}, \tag{13}$$

where we have used the bound $\binom{m_D}{m_U} > (m_D/m_U)^k$ to obtain the estimate on the right hand side. The probability of exact reconstruction using the non-uniform model will always be at least as large as the probability under the uniform model, provided the ratio (13) is one or higher. This holds whenever $m_U \leq m_D s_n \ell_D$. It follows that the probability of exact recovery using the non-uniform measurements is at least the probability of exact recovery from a uniform model with $m_U = \lfloor m_D s_n \ell_D \rfloor$ measurements. This probability is what is given by Theorem 1. $\qquad\square$

B. Noisy Measurements

We now analyze the sensitivity of PhaseLift to the measurement noise $\{\eta_i\}$. For brevity, we focus only on the case of complex-valued signals. To analyze the impact of noise, we re-write the problem (PM) in

the following equivalent form:

$$\begin{cases} \underset{\mathbf{x} \in \mathcal{H}^n}{\text{maximize}} & \langle \mathbf{x}, \hat{\mathbf{x}} \rangle_{\Re} \\ \text{subject to} & |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2 \leq \hat{b}_i^2 + \eta_i, \quad i = 1, 2, \ldots, m. \end{cases} \tag{14}$$

Here, $\hat{b}_i^2 = |\langle \mathbf{a}_i, \mathbf{x}^0 \rangle|^2$ is the (unknown) true magnitude measurement and $b_i^2 = \hat{b}_i^2 + \eta_i$. We are interested in bounding the impact that these measurement errors have on the solution to (PM). Note that the severity of a noise perturbation of size η_i depends on the (arbitrary) magnitude of the measurement vector $\mathbf{a}_i$. For this reason, we assume the vectors $\{\mathbf{a}_i\}$ have unit norm throughout this section.

We will begin by proving results only for the case of non-negative noise. We will then generalize our analysis to the case of arbitrary bounded noise. The following result gives a geometric characterization of the reconstruction error.

Theorem 5. *Suppose the vectors $\{\mathbf{a}_i \in \mathbb{C}^n\}$ in (14) are normalized to have unit length, and the noise vector $\boldsymbol{\eta}$ is non-negative. Let r be the maximum relative noise, defined by*

$$r = \max_{i=1,2,\ldots,m} \left\{ \frac{\eta_i}{b_i} \right\}, \tag{15}$$

and let $\mathcal{D}_{\mathbb{C}} = \{\boldsymbol{\delta} \in \mathcal{S}_{\mathbb{C}}^{n-1} | \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle_{\Re} \geq 0, \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle_{\Im} = 0\}$ be the set of aligned descent directions. Choose some error bound $\varepsilon > r/2$, and define the angle $\theta = \arccos(r/2\varepsilon)$. If the caps $\{\mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i, \theta)\}$ cover $\mathcal{D}_{\mathbb{C}}$, then the solution $\mathbf{x}^\star$ of (PM), and equivalently of the problem in (14), satisfies the bound

$$\|\mathbf{x}^\star - \mathbf{x}^0\|_2 \leq \varepsilon.$$

Proof. We first reformulate the problem (14) as

$$\begin{cases} \underset{\boldsymbol{\Delta} \in \mathbb{C}^n}{\text{maximize}} & \langle \mathbf{x}^0 + \boldsymbol{\Delta}, \hat{\mathbf{x}} \rangle_{\Re} \\ \text{subject to} & |\langle \tilde{\mathbf{a}}_i, (\mathbf{x}^0 + \boldsymbol{\Delta}) \rangle|^2 \leq \hat{b}_i^2 + \eta_i, \quad i = 1, 2, \ldots, m, \end{cases} \tag{16}$$

where $\boldsymbol{\Delta} = \mathbf{x}^\star - \mathbf{x}^0$ is the recovery error vector and $\{\tilde{\mathbf{a}}_i\} = \{\text{phase}(\langle \mathbf{a}_i, \mathbf{x}^0 \rangle) \mathbf{a}_i\}$ are aligned measurement vectors. In this form, the recovery error vector $\boldsymbol{\Delta}$ appears explicitly. Because we assume the errors $\{\eta_i\}$ to be non-negative, the true signal $\mathbf{x}^0$ is feasible for (14). It follows that the optimal objective of the perturbed problem (14) must be at least at large as the optimal value achieved by $\mathbf{x}^0$, i.e., $\langle \boldsymbol{\Delta}, \hat{\mathbf{x}} \rangle_{\Re} \geq 0$. Furthermore, the solution $\mathbf{x}^0 + \boldsymbol{\Delta}$ must be aligned with $\hat{\mathbf{x}}$, as is the true signal $\mathbf{x}^0$, and so $\langle \boldsymbol{\Delta}, \hat{\mathbf{x}} \rangle \in \mathbb{R}$. For the reasons just described, we know that the unit vector $\boldsymbol{\delta} = \boldsymbol{\Delta}/\|\boldsymbol{\Delta}\|_2 \in \mathcal{D}_{\mathbb{C}}$.

Our goal is to put a bound on the magnitude of the recovery error $\boldsymbol{\Delta}$. We start by reformulating the

constraints in (16) to get

$$|\langle \tilde{\mathbf{a}}_i, (\mathbf{x}^0 + \boldsymbol{\Delta}) \rangle|^2 = |\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle|^2 + 2[\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle^* \langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle]_{\Re} + |\langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle|^2 \leq \hat{b}_i^2 + \eta_i.$$

Subtracting $|\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle|^2 = \hat{b}_i^2$ from both sides yields

$$2[\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle^* \langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle]_{\Re} + |\langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle|^2 \leq \eta_i.$$

Since η_i is non-negative, we have

$$\begin{aligned}
\eta_i &\geq 2[\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle^* \langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle]_{\Re} + |\langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle|^2 \\
&= 2\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle \langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle_{\Re} + |\langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle|^2 \\
&\geq 2\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle \langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle_{\Re} + |\langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle_{\Re}|^2.
\end{aligned} \tag{17}$$

The final lower bound can only be less than η_i if

$$\langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta} \rangle_{\Re} \leq -\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle + \sqrt{(\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle)^2 + \eta_i} \leq \frac{\eta_i}{2\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0 \rangle} = \frac{\eta_i}{2\hat{b}_i} \leq \frac{r}{2}. \tag{18}$$

Now suppose that $\|\boldsymbol{\Delta}\|_2 > \varepsilon$. From (18) we have

$$\langle \tilde{\mathbf{a}}_i, \boldsymbol{\delta} \rangle_{\Re} \leq \langle \tilde{\mathbf{a}}_i, \boldsymbol{\Delta}/\|\boldsymbol{\Delta}\|_2 \rangle_{\Re} < \frac{r}{2\varepsilon},$$

Therefore $\boldsymbol{\delta} \notin \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i, \theta)$ where $\theta = \arccos(r/2\varepsilon)$. If the caps $\{C_{\mathbb{C}}(\tilde{\mathbf{a}}_i, \theta)\}$ cover $\mathcal{D}_{\mathbb{C}}$, then $\boldsymbol{\delta} \notin \mathcal{D}_{\mathbb{C}}$, which is a contradiction. It follows that $\|\boldsymbol{\Delta}\|_2 \leq \varepsilon$ if the caps $\{C_{\mathbb{C}}(\tilde{\mathbf{a}}_i, \theta)\}$ cover $\mathcal{D}_{\mathbb{C}}$. $\qquad\square$

Using this result, we can bound the reconstruction error in the noisy case. For brevity, we present results only for the complex-valued case.

Theorem 6. *Suppose the vectors $\{\mathbf{a}_i\}$ in (14) are independenly and uniformly distributed in $\mathcal{S}_{\mathbb{C}}^{n-1}$, and the noise vector $\boldsymbol{\eta}$ is non-negative. Let r be the maximum relative error defined in (15). Choose some error bound $\varepsilon > r/2$, and define the angle $\phi = \arccos(r/2\varepsilon) - \text{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$. Then, the solution $\mathbf{x}^\star$ to (PM) satisfies*

$$\|\mathbf{x}^\star - \mathbf{x}^0\| \leq \varepsilon.$$

with probability at least

$$p_{cover}(m, 2n-1, \phi) \geq 1 - \frac{(em)^{2n-1}\sqrt{2n-2}}{(4n-2)^{2-2}} \exp\left(-\frac{\sin^{2n-2}(\phi)(m-n)}{\sqrt{16n-8}}\right) \cos(\phi) - \exp\left(-\frac{(m-4n+3)^2}{2m-2}\right)$$

when $n \geq 5$ and $m > 4n - 2$.

Proof. Define the following two sets:

$$\mathcal{D} = \{\boldsymbol{\delta} \in \mathcal{S}_{\mathbb{C}}^{n-1} \,|\, \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle \in \mathbb{R}, \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle \geq 0\}$$

$$\mathcal{D}^0 = \{\boldsymbol{\delta} \in \mathcal{S}_{\mathbb{C}}^{n-1} \,|\, \langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle \in \mathbb{R}, \langle \boldsymbol{\delta}, \mathbf{x}^0 \rangle \geq 0\}.$$

We now claim that the conditions of Theorem 5 hold whenever

$$\mathcal{D}^0 \subset \bigcap_i \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i, \phi) \tag{19}$$

where $\{\tilde{\mathbf{a}}_i = \text{phase}(\langle \mathbf{a}_i, \mathbf{x}^0 \rangle)\mathbf{a}_i\}$ is the set of aligned measurement vectors. To prove this claim, choose some $\boldsymbol{\delta} \in \mathcal{D}$ and assume that (19) holds. Since the half-sphere $\mathcal{D}^0$ can be obtained by rotating $\mathcal{D}$ by a principle angle of $\text{angle}(\hat{\mathbf{x}}, \mathbf{x}^0)$, there is some point $\boldsymbol{\delta}^0 \in \mathcal{D}^0$ with $\text{angle}(\boldsymbol{\delta}, \boldsymbol{\delta}^0) \leq \text{angle}(\hat{\mathbf{x}}, \mathbf{x}^0)$. By property (19), there is some cap $\mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i, \phi)$ that contains $\boldsymbol{\delta}_0$. By the triangle inequality for spherical geometry it follows that:

$$\text{angle}(\boldsymbol{\delta}, \tilde{\mathbf{a}}_i) \leq \text{angle}(\boldsymbol{\delta}, \boldsymbol{\delta}^0) + \text{angle}(\boldsymbol{\delta}^0, \tilde{\mathbf{a}}_i) \leq \text{angle}(\mathbf{x}^0, \hat{\mathbf{x}}) + \phi \leq \theta.$$

Therefore, $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i, \theta)$, and the claim is proved.

It only remains to put a bound on the probability that (19) occurs. Note that the aligned vectors $\{\tilde{\mathbf{a}}_i\}$ are uniformly distributed in $\mathcal{D}^0$, which is isomorphic to a half-sphere in $S_{\mathbb{R}}^{2n-2}$. The probability of covering the half sphere $S_{\mathbb{R}}^{2n-2}$ with uniformly distributed caps drawn from that half sphere is at least as great as the probability of covering the whole sphere $S_{\mathbb{R}}^{2n-2}$ with caps drawn uniformly from the entire sphere. This probability is given by Theorem 4 as $p_{\text{cover}}(m, 2n - 1, \phi)$. $\qquad\square$

We now consider the case of non-negative noise. In this case, we bound the error by converting the problem into an equivalent problem with non-negative noise, and then apply Theorem 6.

Theorem 7. *Suppose the vectors $\{\mathbf{a}_i\}$ in (14) are normalized to have unit length. Define the following measures of the noise*

$$s^2 = \min_{i=1,2,\dots,m}\left\{\frac{\hat{b}_i^2 + \eta_i}{\hat{b}_i^2}\right\} = \min_{i=1,2,\dots,m}\left\{\frac{b_i^2}{\hat{b}_i^2}\right\} \quad \text{and} \quad r = \frac{1}{s}\max_{i=1,2,\dots,m}\left\{\hat{b}_i^2 - s^2\hat{b}_i + \frac{\eta_i}{\hat{b}_i}\right\}.$$

Choose some error bound $\varepsilon > r/2$, and define the angle $\phi = \arccos(r/2\varepsilon) - \text{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$. Then, we have the bound

$$\|\mathbf{x}^\star - \mathbf{x}^0\|_2 \leq \varepsilon + (1 - s)\|\mathbf{x}^0\|_2.$$

with probability at least

$$p_{cover}(m, 2n-1, \phi) \geq 1 - \frac{(em)^{2n-1}\sqrt{2n-2}}{(4n-2)^{2-2}} \exp\left(-\frac{\sin^{2n-2}(\phi)(m-n)}{\sqrt{16n-8}}\right)\cos(\phi) - \exp\left(-\frac{(m-4n+3)^2}{2m-2}\right)$$

when $n \geq 5$ and $m > 4n - 2$.

Proof. Consider the "shrunk" version of problem (14)

$$\begin{cases} \underset{\mathbf{x}\in\mathcal{H}^n}{\text{maximize}} & \langle\mathbf{x}, \hat{\mathbf{x}}\rangle_{\Re} \\ \text{subject to} & |\langle\mathbf{a}_i, \mathbf{x}\rangle|^2 \leq s^2\hat{b}_i^2 + \zeta_i, \quad i = 1, 2, \ldots, m. \end{cases} \tag{20}$$

for some real-valued "shrink factor" $s > 0$. Clearly, if $\mathbf{x}^0$ is aligned with $\hat{\mathbf{x}}$ and satisfies $|\langle\mathbf{a}_i, \mathbf{x}^0\rangle| = b_i$ for all i, then $s\mathbf{x}^0$ is aligned with $\hat{\mathbf{x}}$ and satisfies $|\langle\mathbf{a}_i, s\mathbf{x}^0\rangle| = sb_i$. We can now transform the problem (20) into an equivalent problem with non-negative noise by choosing

$$s^2 = \min_{i=1,2,\ldots,m}\left\{\frac{\hat{b}_i^2 + \eta_i}{\hat{b}_i^2}\right\} \quad \text{and} \quad \zeta_i = \hat{b}_i^2 - s^2\hat{b}_i^2 + \eta_i \geq 0.$$

We then have $(sb_i)^2 + \zeta_i = b_i^2 + \eta_i^2$, and so problem (20) is equivalent to problem (14). However, the noise ζ_i in problem (20) is non-negative, and thus we can apply Theorem 6. This theorem requires the constant r for the shrunken problem, which is now

$$r_{\text{shrunk}} = \max_{i=1,2,\ldots,m}\left\{\frac{\zeta_i}{s\hat{b}_i}\right\} = \frac{1}{s}\max_{i=1,2,\ldots,m}\left\{\hat{b}_i^2 - s^2\hat{b}_i + \eta_i/\hat{b}_i\right\}.$$

The solution to the shrunk problem (20) satisfies $\|\mathbf{x}^\star - s\mathbf{x}^0\|_2 \leq \epsilon$, with probability $p_{\text{cover}}(m, 2n - 1, \phi)$, where $\phi = \arccos(r_{\text{shrunk}}/2\epsilon) - \text{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$. If this condition is fulfilled, then we have

$$\|\mathbf{x}^\star - \mathbf{x}^0\|_2 \leq \|\mathbf{x}^\star - s\mathbf{x}^0 + s\mathbf{x}^0 - \mathbf{x}^0\|_2 \leq \|\mathbf{x}^\star - s\mathbf{x}^0\|_2 + \|s\mathbf{x}^0 - \mathbf{x}^0\|_2 \leq \epsilon + (1-s)\|\mathbf{x}^0\|_2,$$

which concludes the proof. $\qquad\square$

VI. How to Compute Approximation Vectors?

There exist a variety of algorithms that compute approximation vectors, such as the (truncated) spectral initializer [20], [23], the Null initializer [25], or the orthogonality-promoting method [24]. We next show that even randomly generated approximation vectors guarantee the success of PhaseMax with high probability given a sufficiently large number of measurements. We then show that more sophisticated methods guarantee success with high probability if the number of measurements depends linearly on n.

A. Random Initialization

Consider the use of approximation vectors $\hat{\mathbf{x}}$ drawn randomly from the unit sphere $S_{\mathbb{R}}^{n-1}$. Do we expect such approximation vectors to be accurate enough to recover the unknown signal? To find out, we analyze the inner product between two real-valued random vectors on the unit sphere. Note that we only care about the *magnitude* of this inner product—if the inner product is negative, then PhaseMax simply recovers $-\mathbf{x}^0$ rather than $\mathbf{x}^0$. Our analysis will make use of the following result.

Lemma 5. *Consider the angle* $\beta = \mathrm{angle}(\mathbf{x}, \mathbf{y})$ *between two random vectors* $\mathbf{x}, \mathbf{y} \in \mathcal{S}_{\mathcal{H}}^{n-1}$ *sampled independently and uniformly from the unit sphere. Then, the expected magnitude of the cosine distance between the two random vectors satisfies*

$$\sqrt{\frac{2}{\pi n}} \leq \mathbb{E}[|\cos(\beta)|] \leq \sqrt{\frac{2}{\pi(n-\frac{1}{2})}}, \quad \textit{for } \mathcal{H} = \mathbb{R} \tag{21}$$

$$\sqrt{\frac{1}{\pi n}} \leq \mathbb{E}[|\cos(\beta)|] \leq \sqrt{\frac{4}{\pi(4n-1)}}, \quad \textit{for } \mathcal{H} = \mathbb{C}. \tag{22}$$

Proof. We first consider the real case. The quantity $\cos(\beta) = \langle \mathbf{x}, \mathbf{y} \rangle / (\|\mathbf{x}\|_2 \|\mathbf{y}\|_2)$ is simply the sample correlation between two random vectors, whose distribution function is given by [32]

$$f(z) = \frac{(1-z^2)^{\frac{n-3}{2}}}{2^{n-2}B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)}, \tag{23}$$

where B is the beta function and $z \in [-1, +1]$. Hence, the expectation of the magnitude of the inner product is given by

$$\mathbb{E}[|\cos(\beta)|] = 2 \frac{\int_0^1 z(1-z^2)^{\frac{n-3}{2}} \, \mathrm{d}z}{2^{n-2}B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)}.$$

The integral in the numerator was studied in [33, Eq. 31] and evaluates to $\frac{1}{2}B(1, \frac{n-1}{2})$. Plugging this expression into (23) and simplifying yields

$$\mathbb{E}[|\cos(\beta)|] = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi}\,\Gamma(\frac{n+1}{2})}.$$

Finally, by using bounds on ratios of Gamma functions [34], we obtain the bounds in (21) for real-valued vectors. The bounds in (22) for complex-valued vectors are obtained by noting that $\mathcal{S}_{\mathbb{C}}^{n-1}$ is isomorphic to $\mathcal{S}_{\mathbb{R}}^{2n-1}$ and by simply replacing $n \leftarrow 2n$ in the bounds for the real-valued case. $\square$

For such randomly-generated approximation vectors, we now consider the approximation accuracy α that appears in Theorems 1 and 3. Note that $\mathbb{E}[|\beta|] \leq \frac{\pi}{2} - \mathbb{E}[|\cos(\beta)|]$, and, thus

$$\mathbb{E}[\alpha] = 1 - \frac{2}{\pi}\mathbb{E}[|\beta|] \geq \frac{2}{\pi}\mathbb{E}[|\cos(\beta)|] \geq \sqrt{\frac{8}{\pi^3 n}}$$

for the real case. Plugging this expected value for α into Theorem 3, we see that, for an average randomly-generated approximation vector, the probability of exact reconstruction goes to 1 rapidly as n goes to infinity, provided that the number of measurements satisfies

$$m > cn^{3/2} \text{ for any } c > \sqrt{\frac{\pi^3}{2}}.$$

For the complex case, Theorem 1 guarantees successful recovery from an average random approximation vector with high probability, provided that

$$m > cn^{3/2} \text{ for any } c > \sqrt{2\pi^3}.$$

Note that such random approximation vectors require $O(n^{3/2})$ measurements rather than the $O(n)$ required by other phase retrieval methods, e.g., [17], [17], [24] (see also Section VII). Hence, it may be more practical for PhaseMax to use approximation vectors obtained from more sophisticated methods.

B. Truncated Spectral Initializer

The truncated spectral initializer [23] is a refinement of the method put forward in [20] and enables the computation of an approximation vector $\hat{\mathbf{x}}$ that exhibits strong theoretical properties. Specifically, the result in [23, Prop. 8] states the following. Fix some $0 < \delta < \sqrt{2}$ and assume that $\|\mathbf{x}^0\|_2 = 1$. Then, with probability exceeding $1 - \exp(-c_0 m)$ for some constant $c_0 > 0$, a unit-length version of the approximation vector $\hat{\mathbf{x}}$ computed by the truncated spectral initializer satisfies $1 - \frac{\delta^2}{2} \leq |\langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle|$, provided that $m > c_1 n$ for some constant $c_1 > 0$. This implies that the approximation accuracy satisfies

$$\alpha = 1 - \tfrac{2}{\pi} \operatorname{angle}(\mathbf{x}^0, \hat{\mathbf{x}}) \geq 1 - \tfrac{2}{\pi} \arccos\!\left(1 - \tfrac{\delta^2}{2}\right) > 0.$$

By combining this result with, for example, Theorem 1, we see that the truncated spectral initializer enables PhaseMax to succeed with high probability provided that $m > c_2 n$ for any constant $c_2 > \max\{4/\alpha, c_1\}$.

VII. DISCUSSION

This section briefly compares our theoretical results to that of existing algorithms. We furthermore demonstrate the sharpness of our recovery guarantees and show the practical limits of PhaseMax.

TABLE I
COMPARISON OF THEORETICAL RECOVERY GUARANTEES FOR NOISELESS PHASE RETRIEVAL

Algorithm	Sample complexity	Lower bound on $p_{\mathbb{C}}(m,n)$
PhaseMax	$m > (4n-1)/\alpha + 1$	$1 - e^{-(\alpha(m-1)-4n+2)^2/(2m-2)}$
PhaseLift [17]	$m \geq c_0 n$	$1 - c_1 e^{-c_2 m}$
TWF [17]	$m \geq c_0 n$	$1 - c_1 e^{-c_2 m}$
TAF [24]	$m \geq c_0 n$	$1 - (m+5)e^{-n/2} - c_1 e^{-c_2 m} - 1/n^2$
Bahmani and Romberg [35]	$m > \frac{32}{\sin^4(\alpha)} \log\left(\frac{8e}{\sin^4(\alpha)}\right)n$	$1 - 8e^{-\sin^4(\alpha)\left(M - \frac{32}{\sin^4(\alpha)} \log\left(\frac{8e}{\sin^4(\alpha)}\right)N\right)/16}$

A. Comparison with Existing Recovery Guarantees

Table I compares our noiseless recovery guarantees in a complex system to that of PhaseLift [17], truncated Wirtinger flow (TWF) [17], and truncated amplitude flow (TAW) [24].[2] We see that PhaseMax requires the same sample complexity (number of required measurements) as compared to PhaseLift, TWF, and TAW, when used together with the truncated spectral initializer [17]. While the constants c_0, c_1, and c_2 in the recovery guarantees for all of the other methods are generally very large, our recovery guarantees contain no unspecified constants, explicitly depend on the approximation factor α, and are surprisingly sharp. We next demonstrate the accuracy of our results via numerical simulations.

After the original version of our manuscript was submitted to a journal, a very recent paper [35] appeared on arXiv proposing an algorithm equivalent to PhaseMax, but with substantially different theoretical results. For completeness, the recovery guarantees from [35] are included in Table I. We emphasize that our recovery guarantees are considerably tighter. For example, in the complex-valued noiseless case with $\text{angle}(\hat{\mathbf{x}}, \mathbf{x}^0) = 45°$, the analysis in [35] requires over $10^5 n$ measurements to guarantee recovery with nonzero probability, whereas our results require just over $8n$ measurements.

B. Accuracy of our Recovery Guarantees

We compare the empirical success probability of PhaseMax in a noiseless and complex-valued scenario with measurement vectors taken independently and uniformly from the unit sphere. We use a custom ADMM-based solver [36] and declare success whenever the relative reconstruction error satisfies

$$RRE = \frac{\|\mathbf{x}^0 - \mathbf{x}\|_2^2}{\|\mathbf{x}^0\|_2^2} < 10^{-5}. \tag{24}$$

---

[2]Since AltMinPhase [20] requires an online measurement model that differs significantly from the other algorithms considered here, we omit a comparison.
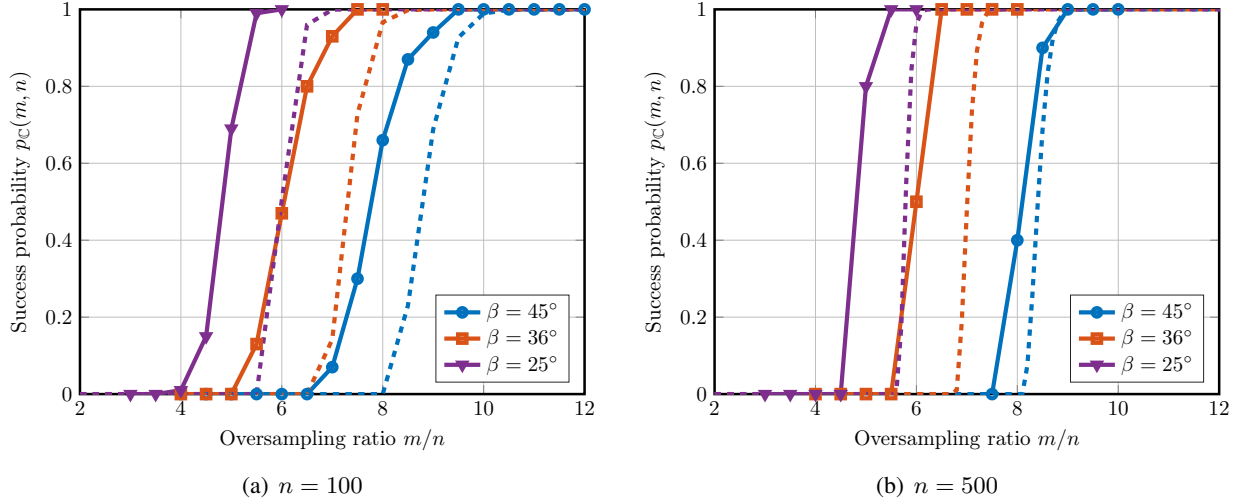
(a) $n = 100$

(b) $n = 500$

Fig. 1. Comparison between the empirical success probability (solid lines) and our theoretical lower bound (dashed lines) for varying angles β between the true signal and the approximation vector. Our theoretical results accurately characterize the empirical success probability of PhaseMax. Furthermore, PhaseMax exhibits a sharp phase transition for larger dimensions.

We compare empirical rates of success to the theoretical lower bound in Theorem 1. Figure 1 shows results for $n = 100$ and $n = 500$ measurements, where we artificially generate an approximation $\hat{\mathbf{x}}$ for different angles $\beta = \text{angle}(\hat{\mathbf{x}}, \mathbf{x}^0)$ measured in degrees. Clearly, our theoretical lower bound accurately predicts the real-world performance of PhaseMax. For large n and large β, the gap between theory and practice becomes extremely tight. We furthermore observe a sharp phase transition between failure and success, with the transition getting progressively sharper for larger dimensions n.

C. Performance Limits of PhaseMax

We briefly compare PhaseMax to a select set of phase retrieval algorithms in terms of the relative reconstruction error. We emphasize that this comparison is by no means intended to be exhaustive and serves the sole purpose of demonstrating the efficacy and limits of PhaseMax (see, e.g., [15], [19] for more extensive phase retrieval algorithm comparisons). We compare the Gerchberg-Saxton algorithm [5], the Fienup algorithm [6], the truncated Wirtinger flow [23], and PhaseMax—all of these methods use the truncated spectral initializer [23]. We also run simulations using the semidefinite relaxation (SDR)-based method PhaseLift [1] implemented via FASTA [37]; this is, together with PhaseCut [19], the only convex alternative to PhaseMax, but lifts the problem to a higher dimension.

Figure 2 reveals that PhaseMax requires larger oversampling ratios m/n to enable faithful signal recovery compared to non-convex phase-retrieval algorithms that operate in the original signal dimension. This is because the truncated spectral initializer requires oversampling ratios of about six or higher to
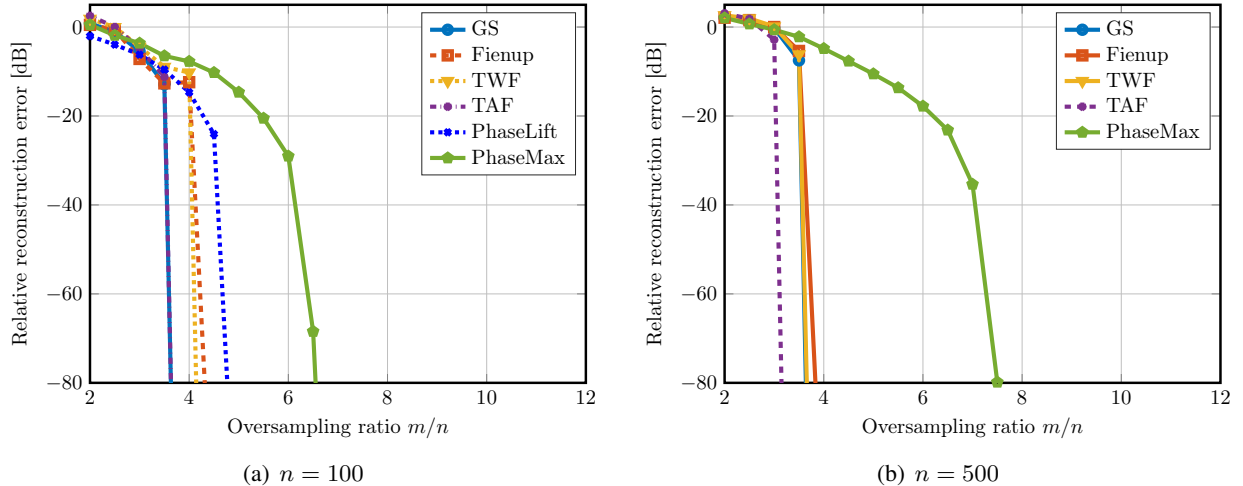
Fig. 2. Comparison of the relative reconstruction error. We use the truncated spectral initializer for Gerchberg-Saxton (GS), Fienup, truncated Wirtinger flow (TWF), truncated amplitude flow (TAF), and PhaseMax. PhaseMax does not achieve exact recovery for the lowest number of measurements among the considered methods, but is convex, operates in the original dimension, and comes with sharp performance guarantees. PhaseLift only terminates in reasonable computation time for $n = 100$.

yield sufficiently accurate approximation vectors $\hat{\mathbf{x}}$ that enable PhaseMax to succeed. While PhaseMax does not achieve exact reconstruction with the lowest number of measurements, it is convex, operates in the original signal dimension, can be implemented via solvers for Basis Pursuit, and comes with *sharp* performance guarantees that do not sweep constants under the rug (cf. Figure 1). The convexity of PhaseMax enables a natural extension to sparse phase retrieval [38], [39] or other signal priors (e.g., total variation or bounded infinity norm) that can be formulated with convex functions. Such non-differentiable priors cannot be efficiently minimized using simple gradient descent methods (which form the basis of Wirtinger or amplitude flow, and many other methods), but can potentially be solved using standard convex solvers when combined with the PhaseMax formulation.

VIII. CONCLUSIONS

We have proposed a novel, convex phase retrieval algorithm, which we call *PhaseMax*. We have provided accurate bounds on the success probability that depend on the signal dimension, the number of measurements, and the angle between the approximation vector and the true vector. Our analysis covers a broad range of random measurement ensembles and characterizes the impact of general measurement noise on the solution accuracy. We have demonstrated the sharpness of our recovery guarantees and studied the practical limits of PhaseMax via simulations.

There are many avenues for future work. We believe that the development of new algorithms that compute more accurate approximation vectors is of significant practical interest, not only for PhaseMax.

Furthermore, extending our results to include useful signal priors (such as the ℓ_1-norm) is an interesting open research topic. Finally, our bounds for the noisy case can be sharpened.

APPENDIX A
PROOF OF LEMMA 4

In this section, we prove Lemma 4. This Lemma is a direct corollary of the following result of Burgisser, Cucker, and Lotz [31]. For a complete proof this result, see Theorem 1.1 of [31], and the upper bound on the constant "C" given in Proposition 5.5.

Theorem 8. *Let $m > n \geq 2$. Then the probability of covering the sphere $S_{\mathbb{R}}^{n-1}$ with independent and uniform random caps of central angle $\phi \leq \pi/2$ is bounded by*

$$p_{cover}(m, n, \phi) \geq 1 - \binom{m}{n} C \int_0^\epsilon (1 - t^2)^{(n^2 - 2n - 1)/2} (1 - \lambda(t))^{m-n} \, dt - \frac{1}{2^{m-1}} \sum_{k=0}^{n-1} \binom{m-1}{k}$$

where $\lambda(t) = \frac{V_{n-1}}{V_n} \int_0^{\arccos(t)} \sin^{n-2}(\phi) \, d\phi$, $V_n = \text{Vol}(S_{\mathbb{R}}^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}$, $C = \frac{n\sqrt{n-1}}{2^{n-1}}$, and $\epsilon = \cos(\phi)$.

While Theorem 8 provides a bound on $p_{\text{cover}}(m, n, \phi)$, the formulation of this bound does not provide any intuition of the scaling of $p_{\text{cover}}(m, n, \phi)$ or its dependence on m and n. For this reason, we derive Lemma 4, which is a weaker but more intuitive result. We restate Lemma 4 here for clarity.

Lemma 4. Let $n \geq 9$, and $m > 2n$. Then the probability of covering the sphere $S_{\mathbb{R}}^{n-1}$ with caps of central angle $\phi \leq \pi/2$ is lower bounded by

$$p_{\text{cover}}(m, n, \phi) \geq 1 - \frac{(em)^n \sqrt{n-1}}{(2n)^{n-1}} \exp\left(-\frac{(1 - \epsilon^2)^{(n-1)/2}(m-n)}{\sqrt{8n}}\right) \cos(\phi) - \exp\left(-\frac{(m - 2n + 1)^2}{2m - 2}\right).$$

Proof. Let us simplify the result of Theorem 8. If we assume $m > 2n$, then Hoeffding's inequality yields

$$\frac{1}{2^{m-1}} \sum_{k=0}^{n-1} \binom{m-1}{k} \leq \exp\left(-\frac{(m - 2n + 1)^2}{2m - 2}\right).$$

Next, we derive a lower bound as follows:

$$\begin{aligned}
\lambda(t) &= \frac{\Gamma(n/2)}{\Gamma((n-1)/2)\sqrt{\pi}} \int_0^{\arccos(t)} \sin^{n-2}(\phi) \, d\phi \\
&\geq \sqrt{(n/2 - 1)/\pi} \int_0^{\arccos(t)} \sin^{n-2}(\phi) \cos(\phi) \, d\phi \\
&= \sqrt{(n/2 - 1)/\pi} \frac{1}{n-1} \sin^{n-1} \arccos(t) \\
&\geq \frac{1}{\sqrt{8n}} (1 - t^2)^{(n-1)/2}.
\end{aligned}$$

We have used the fact that $\sqrt{(n/2-1)/\pi}\frac{1}{n-1} > \frac{1}{\sqrt{8n}}$ for $n \geq 4$, and also the "Wallis ratio" bound $\frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \geq \sqrt{n/2-1}$ [40], [41]. Finally, we plug in the inequality $\binom{m}{n} \leq \frac{(em)^n}{n^n}$. We now have

$$\binom{m}{n} C \int_0^\epsilon (1-t^2)^{(n^2-2n-1)/2}(1-\lambda(t))^{m-n}\, \mathrm{d}t$$

$$\leq \frac{(em)^n \sqrt{n-1}}{(2n)^{n-1}} \int_0^\epsilon (1-t^2)^{(n^2-2n-1)/2}\left(1 - \frac{1}{\sqrt{8n}}(1-t^2)^{(n-1)/2}\right)^{m-n}\, \mathrm{d}t. \quad (25)$$

Now we simplify the integral. Using the identity $(1-x)^a < e^{-ax}$, which holds for $x \leq 1$, we can convert each term in the integrand into an exponential. We do this first with $\mathbf{x} = t^2$ and then with $\mathbf{x} = \frac{1}{\sqrt{8n}}(1-t^2)^{(n-1)/2}$ to obtain

$$(1-t^2)^{(n^2-2n-1)/2}\left(1 - \frac{1}{\sqrt{8n}}(1-t^2)^{(n-1)/2}\right)^{m-n}$$

$$\leq \exp\left(-\frac{t^2(n^2-2n-1)}{2} - \frac{(1-t^2)^{(n-1)/2}(m-n)}{\sqrt{8n}}\right). \quad (26)$$

We then apply the Cauchy-Schwarz inequality to get

$$\int_0^\epsilon \exp\left(-\frac{t^2(n^2-2n-1)}{2} - \frac{(1-t^2)^{(n-1)/2}(m-n)}{\sqrt{8n}}\right)\, \mathrm{d}t$$

$$\leq \left[\int_0^\epsilon \exp\left(-t^2(n^2-2n-1)\right)\, \mathrm{d}t\right]^{1/2}\left[\int_0^\epsilon \exp\left(-\frac{(1-t^2)^{(n-1)/2}(m-n)}{\sqrt{2n}}\right)\, \mathrm{d}t\right]^{1/2}$$

$$\leq [\epsilon]^{1/2}\left[\epsilon\exp\left(-\frac{(1-\epsilon^2)^{(n-1)/2}(m-n)}{\sqrt{2n}}\right)\right]^{1/2} = \epsilon\exp\left(-\frac{(1-\epsilon^2)^{(n-1)/2}(m-n)}{\sqrt{8n}}\right).$$

Replacing the integral with this bound yields the result.

$\square$

REFERENCES

[1] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jul. 2001.

[3] S. Chen and D. Donoho, "Basis pursuit," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 1, Oct. 1994, pp. 41–44.

[4] C. Studer, W. Yin, and R. G. Baraniuk, "Signal representations with minimum ℓ_∞-norm," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Oct. 2012, pp. 1270–1277.

[5] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, Aug. 1972.

[6] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.

[7] R. W. Harrison, "Phase problem in crystallography," *J. Opt. Soc. Am. A*, vol. 10, no. 5, pp. 1046–1055, May 1993.

[8] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, "Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes," *Ann. Rev. Phys. Chem.*, vol. 59, pp. 387–410, Nov. 2008.

[9] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nat. Phys.*, vol. 2, no. 4, pp. 258–261, Apr. 2006.

[10] S. S. Kou, L. Waller, G. Barbastathis, and C. J. Sheppard, "Transport-of-intensity approach to differential interference contrast (TI-DIC) microscopy for quantitative phase imaging," *Opt. Lett.*, vol. 35, no. 3, pp. 447–449, Feb. 2010.

[11] H. Faulkner and J. Rodenburg, "Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm," *Phys. Rev. Lett.*, vol. 93, no. 2, p. 023903, Jul. 2004.

[12] J. Holloway, M. S. Asif, M. K. Sharma, N. Matsuda, R. Horstmeyer, O. Cossairt, and A. Veeraraghavan, "Toward long-distance subdiffraction imaging using coherent camera arrays," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 251–265, Sept. 2016.

[13] F. Fogel, I. Waldspurger, and A. d'Aspremont, "Phase retrieval for imaging problems," *Math. Prog. Comp.*, vol. 8, no. 3, pp. 311–335, Sept. 2016.

[14] E. J. Candès, E.es, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Appl. Comput. Harm.. Anal.*, vol. 39, no. 2, pp. 277–299, Sept. 2015.

[15] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," *arXiv:1510.07713*, Oct. 2015.

[16] L. Tian and L. Waller, "3D intensity and phase imaging from light field measurements in an LED array microscope," *Optica*, vol. 2, no. 2, pp. 104–111, Feb. 2015.

[17] E. J. Candès and X. Li, "Solving quadratic equations via phaselift when there are about as many equations as unknowns," *Found. Comput. Math.*, vol. 14, no. 5, pp. 1017–1026, Oct. 2014.

[18] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Rev.*, vol. 57, no. 2, pp. 225–251, Nov 2015.

[19] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Math. Prog.*, vol. 149, no. 1-2, pp. 47–81, Feb. 2015.

[20] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Adv. Neural Inf. Process. Syst.*, 2013, pp. 2796–2804.

[21] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Sig. Process.*, vol. 63, no. 4, pp. 1043–1055, Feb. 2015.

[22] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Feb. 2015.

[23] Y. Chen and E. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Adv. Neural Inf. Process. Syst.*, 2015, pp. 739–747.

[24] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *arXiv: 1605.08285*, Jul. 2016.

[25] P. Chen, A. Fannjiang, and G.-R. Liu, "Phase retrieval with one or two diffraction patterns by alternating projections of the null vector," *arXiv:1510.07379*, Apr. 2015.

[26] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," *arXiv:1602.06664*, Mar. 2016.

[27] L. Schläfli, *Gesammelte Mathematische Abhandlungen I.* Springer Basel, 1953.

[28] J. G. Wendel, "A problem in geometric probability," *Math. Scand.*, vol. 11, pp. 109–111, 1962.

[29] E. Gilbert, "The probability of covering a sphere with n circular caps," *Biometrika*, vol. 52, no. 3/4, pp. 323–330, Dec. 1965.

[30] Z. Füredi, "Random polytopes in the d-dimensional cube," *Disc. Comput. Geom.*, vol. 1, no. 4, pp. 315–319, Dec. 1986.

[31] P. Bürgisser, F. Cucker, and M. Lotz, "Coverage processes on spheres and condition numbers for linear programming," *Ann. Probab.*, vol. 38, no. 2, pp. 570–604, 2010.

[32] J. F. Kenney and E. Keeping, *Mathematics of Statistics, Part 2*. D. Van Nostrand, 1951.

[33] L. Jacques, "A quantized Johnson–Lindenstrauss lemma: The finding of Buffon's needle," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 5012–5027, Sept. 2015.

[34] F. Qi and Q.-M. Luo, "Bounds for the ratio of two gamma functions—from Wendel's and related inequalities to logarithmically completely monotonic functions," *Banach J. Math. Anal*, vol. 6, no. 2, pp. 132–158, May. 2012.

[35] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," *arXiv:1610.04210*, Oct. 2016.

[36] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM J. Imag. Sci.*, vol. 7, no. 3, pp. 1588–1623, 2014.

[37] T. Goldstein, C. Studer, and R. Baraniuk, "A field guide to forward-backward splitting with a FASTA implementation," *arXiv:1411.3406*, Feb. 2014.

[38] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Convex algorithms and limitations," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 1022–1026.

[39] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: efficient phase retrieval of sparse signals," *IEEE Trans. Sig. Process.*, vol. 62, no. 4, pp. 928–938, Jan. 2014.

[40] C. Mortici, "New approximation formulas for evaluating the ratio of gamma functions," *Math. Comput. Model.*, vol. 52, no. 1, pp. 425–433, Jul. 2010.

[41] W. Gautschi, "Some elementary inequalities relating to the gamma and incomplete gamma function," *J. Math. Phys.*, vol. 38, no. 1, pp. 77–81, Apr. 1959.